

KETEPATAN KLASIFIKASI STATUS KERJA DI KOTA TEGAL MENGUNAKAN ALGORITMA C4.5 DAN *FUZZY K-NEAREST NEIGHBOR IN EVERY CLASS* (FK-NNC)

Atika Elsadining Tyas¹, Dwi Ispriyanti², Sudarno³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

Email: tika.elsa43@gmail.com ispriyanti.dwi@gmail.com dsghani@gmail.com

ABSTRACT

Unemployment is a very crucial problem that always deal a developing country and affected a national foundation. It used two methods for classifying a employment status on productive society in Tegal City on August 2014, the methods are C4.5 Algorithm and *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). C4.5 Algorithm is a way of classifying methods from data mining that use to construct a decision tree. FK-NNC is another classification technique that predict using the amount of closest neighbor of K in every class from a testing data. The predictor variables that used on classifying an employment status are neighborhood status, sex, age, marriage status, education, and a work training. To evaluate the result of classification use APER calculation. Based on this analysis, classification of employment status using C4.5 Algorithm obtained APER = 28,3784% and 71,6216% of accuracy, while FK-NNC methods obtained APER = 21,62% and 78,38% of accuracy. So, it can be concluded that FK-NNC is better than C4.5 Algorithm.

Keywords: Classification, C4.5 Algorithm, *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC), APER

1. PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara kepulauan yang memiliki jumlah penduduk yang besar, serta dikategorikan sebagai negara berkembang. Indonesia sebagai negara berkembang perlu melakukan banyak perubahan untuk mendukung pembangunan nasional. Pembangunan nasional bertujuan untuk meningkatkan kesejahteraan masyarakat. Menurut BPS (2014), aspek-aspek yang dapat menggambarkan kondisi tingkat kesejahteraan rakyat pada saat ini meliputi kependudukan, kesehatan dan gizi, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, perumahan dan lingkungan, kemiskinan, serta sosial lainnya.

Salah satu masalah yang menjadi perhatian pemerintah adalah permasalahan di bidang ketenagakerjaan terutama masalah pengangguran. Masalah pengangguran merupakan masalah yang sangat krusial yang selalu dihadapi negara berkembang dan dapat mempengaruhi pembangunan nasional. Tingginya pengangguran di suatu negara mengakibatkan rendahnya partisipasi rakyat dalam pertumbuhan ekonomi yaitu pendapatan rakyat akan berkurang sehingga dapat menyebabkan timbulnya kemiskinan dan masalah sosial lainnya. Dari data Badan Pusat Statistik (BPS) diketahui angkatan kerja Indonesia pada Agustus 2014 mencapai angka 121,9 juta orang. Sedangkan, penduduk yang bekerja pada Agustus 2014 adalah sebanyak 114,6 juta orang. Untuk angka Tingkat Pengangguran Terbuka (TPT) pada Agustus 2014, menurut Kepala BPS Suryamin adalah sebesar 5,94%. Jika dibandingkan dengan Agustus 2013, tingkat pengangguran terbuka mengalami penurunan dari 6,17% menjadi 5,94%. Berdasarkan permasalahan di atas, perlu dilakukan pengklasifikasian status kerja bahwa seseorang dikatakan pengangguran atau bukan pengangguran.

Pada penelitian ini digunakan dua metode untuk mengklasifikasikan data status kerja tersebut, yaitu Algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC),

dimana kedua metode tersebut merupakan teknik klasifikasi yang sederhana tetapi mempunyai hasil kerja yang cukup bagus. Algoritma C4.5 adalah salah satu metode klasifikasi dari *data mining* yang digunakan untuk mengkonstruksikan pohon keputusan (*decision tree*). *Fuzzy K-Nearest Neighbor in every Class* atau dapat disingkat FK-NNC adalah teknik klasifikasi yang melakukan prediksi menggunakan sejumlah K tetangga terdekat pada setiap kelas dari sebuah data uji. Penelitian sebelumnya pernah dilakukan oleh Eko Prasetyo (2012b) pada klasifikasi *data set* iris dan *vertebral column* dengan membandingkan tiga metode yaitu metode *K-Nearest Neighbor* (K-NN), *Fuzzy K-Nearest Neighbor* (FK-NN), dan *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). Pada penelitian tersebut didapatkan kesimpulan bahwa nilai akurasi yang diberikan oleh FK-NNC terbukti lebih tinggi daripada K-NN dan FK-NN.

Pada penelitian ini akan diteliti mengenai data status kerja Kota Tegal tahun 2014 menggunakan Algoritma C4.5 yang akan dibandingkan dengan menggunakan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). Dari kedua metode tersebut masing-masing akan menghasilkan nilai ketepatan klasifikasi. Sehingga, dapat dicari ketepatan klasifikasi yang terbaik dari kedua metode tersebut.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah membentuk pohon klasifikasi menggunakan Algoritma C4.5, mengukur ketepatan klasifikasi menggunakan Algoritma C4.5 dan FK-NNC, serta membandingkan hasil ketepatan klasifikasi status kerja untuk Kota Tegal tahun 2014 menggunakan kedua metode.

2. TINJAUAN PUSTAKA

2.1 Status Kerja

Penduduk yang tergolong dalam angkatan kerja terbagi menjadi dua kategori, yaitu bekerja (*Employment*) dan pengangguran (*Unemployment*). Sedangkan pengangguran sendiri terbagi menjadi dua macam yaitu pengangguran terbuka dan setengah pengangguran (BPS, 2008). Status kerja dalam penelitian ini digunakan status kerja untuk data pengangguran terbuka. Pengangguran terbuka didefinisikan sebagai orang yang sedang mencari pekerjaan atau yang sedang mempersiapkan usaha atau juga yang tidak mencari pekerjaan karena merasa tidak mungkin lagi mendapat pekerjaan, termasuk juga mereka yang baru mendapat kerja tetapi belum mulai bekerja. Pengangguran terbuka tidak termasuk orang yang masih sekolah atau mengurus rumah tangga, sehingga hanya orang yang termasuk angkatan kerja saja yang merupakan pengangguran terbuka (BPS, 2013).

2.2 Data Mining

Data mining, sering juga disebut *Knowledge Discovery in Database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Santosa, 2007).

2.3 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang dapat dipakai dalam pembentukan pohon keputusan. Menurut Widodo *dkk.*, (2013), pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam data *mining* dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus. Tugas paling umum yang diserahkan kepada pohon keputusan adalah klasifikasi. Algoritma C4.5 diperkenalkan oleh

Quinlan (1993) sebagai versi perbaikan dari Algoritma *Iterative Dichotomiser 3* (ID3). Menurut Witten *et al.*, (2011), Algoritma C4.5 memiliki keunggulan dibandingkan dengan ID3 yaitu mampu mengatasi nilai yang hilang (*missing value*), mengatasi data bertipe kontinu, dan melakukan pemangkasan pohon (*prunning trees*).

Prosedur Pemilahan Algoritma C4.5

Algoritma C4.5 menggunakan kriteria *gain* dalam menentukan pemecah *node* pada pohon keputusan. Menurut Rokach dan Maimon (2008), *information gain* atau yang bisa disebut dengan *gain info* adalah kriteria pemisahan yang menggunakan pengukuran *entropy*. *Entropy* adalah rata-rata jumlah informasi yang dibutuhkan untuk mengidentifikasi kelas pada kasus ke dalam himpunan T . Nilai dari setiap penghitungan *entropy* memiliki satuan *bits* atau *binary digits* (Quinlan, 1993). *Entropy* digunakan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dari suatu kumpulan sampel data. Jika kumpulan sampel data semakin heterogen, maka nilai *entropy*-nya semakin besar.

Menurut Ruggieri (2002), *information gain* atribut a dari suatu himpunan T dapat dihitung sebagai berikut. Jika sebuah atribut a adalah diskret dari suatu himpunan kasus T dan T_1, \dots, T_s adalah sub-himpunan dari T yang terdiri dari kasus-kasus yang nilainya sudah diketahui maka untuk mendapatkan *information gain* dari atribut a atau $Gain(a)$ dibutuhkan *entropy* keseluruhan kelas atau $info(T)$ dan *entropy* masing-masing atribut pada himpunan T atau $info(T_i)$. Rumus dari $Gain(a)$ adalah sebagai berikut:

$$Gain(a) = info(T) - \sum_{i=1}^s \frac{|T_i|}{|T|} \times info(T_i) \quad (1)$$

dimana nilai *entropy* keseluruhan kelas:

$$info(T) = - \sum_{j=1}^n \frac{freq(C_j, T)}{|T|} \times {}^2 \log \left(\frac{freq(C_j, T)}{|T|} \right) \quad (2)$$

sedangkan nilai *entropy* untuk setiap atribut i :

$$info(T_i) = - \sum_{j=1}^n \frac{freq(C_j, T_i)}{|T_i|} \times {}^2 \log \left(\frac{freq(C_j, T_i)}{|T_i|} \right) \quad (3)$$

keterangan:

| | |
|----------------|--|
| $ T $ | = Banyaknya kasus dalam himpunan T |
| $ T_i $ | = Banyaknya kasus dalam sub-himpunan T_i |
| $freq(C_j, T)$ | = Banyak dari kasus-kasus dalam himpunan T yang memiliki kelas C_j |

Jika a adalah atribut kontinu maka kasus dalam T dengan nilai atribut tersebut diurutkan dari yang terkecil sampai terbesar. Dimisalkan nilai hasil pengurutan adalah w_1, \dots, w_m , dan nilai $v = \frac{(w_i + w_{i+1})}{2}$ dimana $i \in [1, m-1]$ dan pemisahan yang terjadi untuk atribut bertipe kontinu adalah:

$$T_1^v = \{w_j | w_j \leq v\} \text{ dan } T_2^v = \{w_j | w_j > v\} \quad (4)$$

Untuk setiap nilai v , *gain info* dari *gain* dihitung dengan mempertimbangkan prosedur pemisahan (4). *Information gain* untuk a didefinisikan sebagai nilai maksimum dari semua *gain* dan nilai v merupakan sebagai nilai ambang batas untuk atribut kontinu.

2.4 Fuzzy K-Nearest Neighbor in every Class

Menurut Prasetyo (2012a), metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC) menggunakan sejumlah K tetangga terdekat pada setiap kelas dari sebuah data uji. Kerangka kerja FK-NNC menggunakan FK-NN sebagai basis kerangka kerja, dimana sebuah data uji mempunyai nilai keanggotaan pada setiap kelas dalam interval $[0,1]$. Jumlah nilai keanggotaan sebuah data pada semua kelas sama dengan 1, seperti pada persamaan berikut:

$$\sum_{k=1}^c u_{ik} = 1, \quad 0 \leq u_{ik} \leq 1 \quad (5)$$

dimana u_{ik} adalah nilai keanggotaan data uji x_i ke kelas- k .

Setiap data uji x_i , harus dicari K tetangga terdekat pada setiap kelas. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean. Jarak Euclidean memberikan jarak lurus antara dua buah data dengan N dimensi. Rumus jarak Euclidean adalah sebagai berikut (Han and Kamber, 2006):

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^N (\text{diff}(x_{il}, x_{jl}))^2} \quad (6)$$

dengan :

- x_{il} = data testing ke- i pada variabel ke- l
- x_{jl} = data training ke- j pada variabel ke- l
- $d(x_i, x_j)$ = jarak
- N = dimensi data variabel bebas
- $\text{diff}(x_{il}, x_{jl})$ = *difference* atau selisih

Penghitungan nilai *difference* atau selisih pada persamaan (6) tergantung pada tipe data yang digunakan. Menurut Prasetyo (2012a), penghitungan nilai selisih berdasarkan tipe data untuk tiap variabel dapat diringkas seperti pada Tabel 1.

Tabel 1. Formula Jarak Dua Data dengan Satu Atribut

| Tipe Atribut | Formula Jarak |
|---------------------|--|
| Nominal | $d_{ij} = \begin{cases} 0 & \text{Jika } x_i = x_j \\ 1 & \text{Jika } x_i \neq x_j \end{cases}$ |
| Ordinal | $d_{ij} = x_i - x_j / (n - 1)$ n adalah jumlah pengkategorian dalam x |
| Interval atau Rasio | $d_{ij} = x_i - x_j $ |

Jarak data uji x_i ke semua K tetangga dari setiap kelas ke- k dijumlahkan. Rumus yang digunakan adalah sebagai berikut:

$$S_{ik} = \sum_{j=1}^K d(x_i, x_j)^{\frac{-2}{m-1}} \quad (7)$$

Nilai d sebagai akumulasi jarak data uji x_i ke K tetangga dalam kelas ke- k dilakukan sebanyak C kelas. Nilai m disini merupakan pangkat bobot (*weight exponent*). Selanjutnya, akumulasi jarak data uji x_i ke setiap kelas digabungkan, disimbolkan D . Rumus yang digunakan adalah sebagai berikut:

$$D_i = \sum_{k=1}^C (S_{ik}) \quad (8)$$

Untuk mendapatkan nilai keanggotaan data uji x_i pada setiap kelas ke- k (ada C kelas) dapat digunakan rumus berikut:

$$u_{ik} = \frac{S_{ik}}{D_i} \quad (9)$$

Untuk menentukan kelas hasil prediksi data uji x_i , dipilih kelas dengan nilai keanggotaan terbesar dari data x_i . Rumus yang digunakan adalah sebagai berikut:

$$y' = \operatorname{argmax} \{u_{ik}\}_{k=1,2,\dots,C} \quad (10)$$

dengan:

$$\begin{aligned} y' &= \text{kelas prediksi} \\ k &= 1, 2, \dots, C \\ C &= \text{banyak kelas} \end{aligned}$$

2.5 Ketepatan Klasifikasi

APER (*Apparent Error Rate*) atau yang disebut laju error merupakan ukuran evaluasi yang digunakan untuk melihat peluang kesalahan klasifikasi yang dihasilkan oleh suatu fungsi klasifikasi. Nilai APER menunjukkan proporsi observasi yang salah diklasifikasikan oleh fungsi klasifikasi. Semakin kecil nilai APER maka hasil pengklasifikasian semakin baik (Prasetyo, 2012a).

Menurut Johnson dan Wichern (2007), terjadinya kesalahan klasifikasi suatu observasi merupakan hal yang sangat mungkin terjadi. Hal ini dikarenakan terkadang terdapat beberapa observasi yang tidak berasal dari kelompok tertentu tetapi dimasukkan ke dalam kelompok tersebut. Perhitungan nilai *Apparent Error Rates* (APER) dapat dilakukan dengan menggunakan matriks konfusi sebagai berikut:

Tabel 2. Matrik Konfusi

| Kelompok Aktual | Kelompok Prediksi | | Jumlah Observasi |
|--------------------|----------------------|----------|---------------------|
| | 1 | 2 | |
| 1 | n_{11} | n_{12} | n_1 |
| 2 | n_{21} | n_{22} | n_2 |

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (5)$$

dimana:

$$\begin{aligned} n_{11} &= \text{banyak data dalam kelompok 1 yang secara benar dipetakan ke kelompok 1} \\ n_{12} &= \text{banyak data dalam kelompok 1 yang dipetakan secara salah ke kelompok 2} \\ n_{22} &= \text{banyak data dalam kelompok 2 yang secara benar dipetakan ke kelompok 2} \\ n_{21} &= \text{banyak data dalam kelompok 2 yang dipetakan secara salah ke kelompok 1} \end{aligned}$$

Dari perhitungan nilai APER yang telah diuraikan tersebut, maka dapat dilihat nilai errornya. Sehingga untuk mencari nilai ketepatannya dapat menggunakan $1 - APER$.

3. METODOLOGI PENELITIAN

3.1 Sumber Data

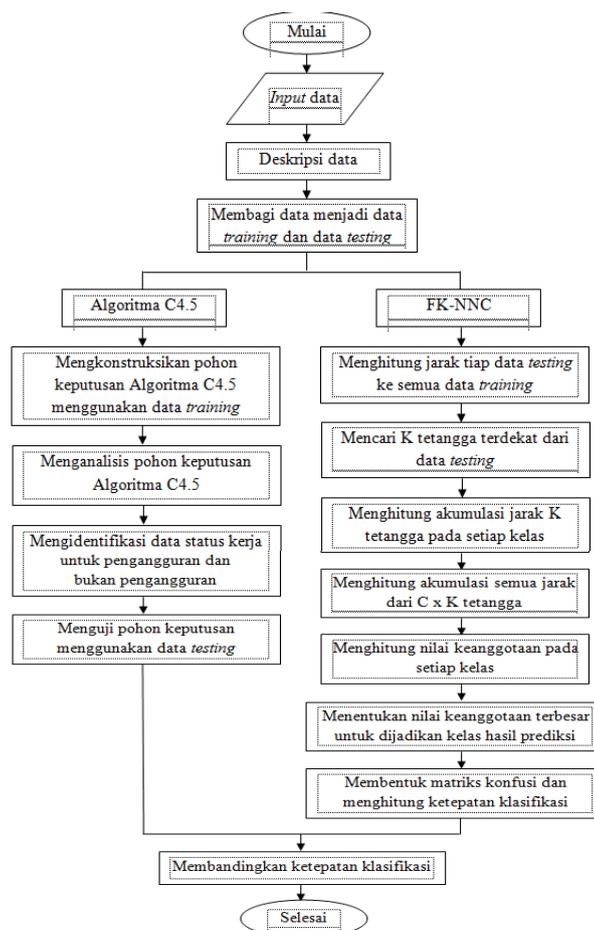
Data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut merupakan data pengangguran terbuka yang diambil dari hasil Survei Angkatan Kerja Nasional (SAKERNAS) untuk Kota Tegal oleh Badan Pusat Statistik pada bulan Agustus 2014 dimana objek penelitiannya berupa data rumah tangga.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah status kerja (pengangguran dan bukan pengangguran), status dalam rumah tangga (kepala rumah tangga dan bukan kepala rumah tangga), jenis kelamin (laki-laki dan perempuan), umur, status perkawinan (belum kawin, kawin, dan cerai), pendidikan (rendah, sedang, dan tinggi), dan pelatihan kerja (pernah dan tidak pernah).

3.3 Metode Analisis

Data sekunder yang telah terkumpul kemudian dianalisis dan diolah dengan menggunakan Algoritma C4.5 dan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC) dengan menggunakan *software* WEKA (*Waikato Environment for Knowledge Analysis*), Matlab R2009a, dan Microsoft Excel. Adapun diagram alir pengolahan data menggunakan Algoritma C4.5 dan metode FK-NNC sebagai berikut:



Gambar 1. Diagram Alir Pengolahan Data Menggunakan Algoritma C4.5 dan FK-NNC

4. HASIL DAN PEMBAHASAN

4.1 Status Kerja Kota Tegal Bulan Agustus 2014

Berdasarkan data Survei Angkatan Kerja Nasional (SAKERNAS) untuk status kerja untuk Kota Tegal pada bulan Agustus 2014 diperoleh informasi seperti berikut:

Tabel 3. Status Kerja Kota Tegal Bulan Agustus 2014

| Status Kerja | Jumlah | Persentase |
|--------------------|--------|------------|
| Pengangguran | 59 | 32% |
| Bukan Pengangguran | 126 | 68% |
| Total | 185 | 100% |

Berdasarkan Tabel 3 dapat diketahui bahwa jumlah penduduk angkatan kerja di Kota Tegal yang berstatus bukan pengangguran lebih besar daripada jumlah penduduk yang berstatus pengangguran. Jumlah penduduk yang berstatus bukan pengangguran sebesar 68% atau 126 orang sedangkan jumlah penduduk yang berstatus pengangguran sebesar 32% atau 59 orang.

4.2 Pengklasifikasian Menggunakan Algoritma C4.5

4.2.1 Konstruksi Algoritma C4.5

Untuk membagi data *training* dan *testing* dilakukan beberapa kali percobaan dengan melihat hasil akurasi yang paling tinggi. Pada penelitian ini, data dipartisi 60% untuk data *training* atau sebanyak 111 data dan 40% data *testing* atau sebanyak 74 data. Konstruksi pohon keputusan menghasilkan 24 simpul, yang terdiri dari simpul akar, simpul keputusan, dan simpul daun. Banyak simpul daun sendiri mencapai 15 simpul yang dilabeli dengan kelasnya masing-masing. Simpul akar merupakan simpul yang terletak paling atas. Berikut ini adalah hasil perhitungan mencari nilai *entropy* dan *information gain* dari semua atribut untuk menentukan simpul akar:

Tabel 4. Nilai *Information Gain* pada Simpul Akar

| No | Atribut | <i>Gain(a)</i> (dalam bits) |
|----------|---------------------------|--------------------------------|
| 1 | Status dalam Rumah Tangga | 0,03696 |
| 2 | Jenis Kelamin | 0,04149 |
| 3 | Umur | 0 |
| 4 | Status Perkawinan | 0,01318 |
| 5 | Pendidikan | 0,00965 |
| 6 | Pelatihan Kerja | 0,01502 |

Berdasarkan Tabel 4 dapat diketahui bahwa dalam penelitian ini atribut jenis kelamin terpilih sebagai pemilah pada simpul akar karena memiliki nilai *information gain* terbesar diantara atribut lainnya.

4.2.2 Pengujian Hasil Pohon Keputusan

Hasil konstruksi pohon baik digunakan untuk memprediksi kemungkinan kelas pada kasus-kasus selanjutnya, pohon konstruksi Algoritma C4.5 tersebut diujikan dengan memasukkan data *testing* ke dalam pohon konstruksi. Tabel matriks konfusinya sebagai berikut:

Tabel 5. Matriks Konfusi Data *Testing*

| Kelompok Aktual | Kelompok Prediksi | | Total |
|--------------------|-------------------|----|-------|
| | 1 | 2 | |
| 1 | 6 | 18 | 24 |
| 2 | 3 | 47 | 50 |
| Total | 9 | 65 | 74 |

Berdasarkan Tabel 5 dapat diketahui bahwa pada pengklasifikasian data status kerja Kota Tegal Bulan Agustus 2014 didapatkan nilai APER sebesar 0,283784 atau 28,3784 dan nilai ketepatan klasifikasi sebesar 0,716216 atau 71,6216%.

4.3 Pengklasifikasian Menggunakan Metode FK-NNC

Metode FK-NNC pada penelitian ini menggunakan kombinasi data *training* 60% atau sebanyak 111 data dan data *testing* 40% atau sebanyak 74 data. Data *training* dan *testing* diolah menggunakan metode FK-NNC dengan nilai K berturut-turut adalah 1, 2, 3, 4, 5, 6, 7, 8, 9. Sehingga dihasilkan nilai ketepatan klasifikasinya sebagai berikut:

Tabel 6. Hasil Ketepatan Klasifikasi Menggunakan FK-NNC

| K | Ketepatan Klasifikasi FK-NNC | APER |
|----------|------------------------------|---------------|
| 1 | 63,51% | 36,49% |
| 2 | 75,68% | 24,32% |
| 3 | 78,38% | 21,62% |
| 4 | 77,03% | 22,97% |
| 5 | 74,32% | 25,68% |
| 6 | 74,32% | 25,68% |
| 7 | 72,97% | 27,03% |
| 8 | 72,97% | 27,03% |
| 9 | 72,97% | 27,03% |

Berdasarkan Tabel 6 dapat diketahui tentang pengolahan data menggunakan FK-NNC dengan kombinasi data *training* dan data *testing* sebesar 60% : 40%. Hasil yang diperoleh untuk $K = 3$ mempunyai nilai ketepatan klasifikasi tertinggi dibandingkan nilai ketepatan klasifikasi untuk $K = 1, 2, 4, 5, 6, 7, 8,$ dan 9. Nilai ketepatan klasifikasi yang diperoleh untuk $K = 3$ sebesar 78,38% dan APER sebesar 21,62%.

4.4 Perbandingan Ketepatan Klasifikasi

Ketepatan klasifikasi menggunakan Algoritma C4.5 dan FK-NNC dapat dilihat dalam Tabel 7.

Tabel 7. Perbandingan Ketepatan Klasifikasi

| Metode | Ketepatan Klasifikasi | APER |
|----------------|-----------------------|----------|
| Algoritma C4.5 | 71,6216% | 28,3784% |
| FK-NNC | 78,38% | 21,62% |

Berdasarkan Tabel 7 diperoleh nilai ketepatan klasifikasi dengan menggunakan Algoritma C4.5 sebesar 71,6216% dan APER = 28,3784%. Sedangkan nilai ketepatan klasifikasi menggunakan FK-NNC sebesar 78,38% dan APER = 21,62%. Nilai ketepatan klasifikasi yang dihasilkan oleh FK-NNC lebih tinggi daripada nilai ketepatan klasifikasi yang dihasilkan oleh Algoritma C4.5 dan laju error (APER) yang dihasilkan oleh FK-NNC lebih kecil daripada laju error (APER) yang dihasilkan oleh Algoritma C4.5. Sehingga dapat disimpulkan bahwa metode FK-NNC lebih baik dalam mengklasifikasikan status kerja untuk Kota Tegal.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dipaparkan pada bab sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

- 1) Konstruksi pohon keputusan yang terbentuk menggunakan Algoritma C4.5 menghasilkan pohon dengan banyak simpul mencapai 24 simpul dimana 15 diantaranya adalah simpul daun dan atribut jenis kelamin terpilih sebagai simpul akar dalam pembuatan pohon keputusan.
- 2) Berdasarkan hasil analisis menggunakan Algoritma C4.5 didapatkan ketepatan klasifikasi status kerja sebesar 71,6216% dan APER = 28,3784%.
- 3) Berdasarkan hasil analisis menggunakan FK-NNC menunjukkan bahwa pada $K = 3$ diperoleh ketepatan klasifikasi status kerja sebesar 78,38% dan APER = 21,62%.
- 4) Nilai ketepatan klasifikasi yang dihasilkan oleh FK-NNC lebih tinggi daripada nilai ketepatan klasifikasi yang dihasilkan oleh Algoritma C4.5 dan laju error (APER) yang dihasilkan oleh FK-NNC lebih kecil daripada laju error (APER) yang dihasilkan oleh Algoritma C4.5. Sehingga dapat disimpulkan bahwa metode FK-NNC lebih baik dalam mengklasifikasikan status kerja untuk Kota Tegal.

DAFTAR PUSTAKA

- BPS Provinsi Jawa Tengah. 2008. *Profil Ketenagakerjaan Jawa Tengah Hasil Sakernas Februari 2008*. Semarang: Badan Pusat Statistik Provinsi Jawa Tengah.
- BPS Provinsi Jawa Tengah. 2013. *Keadaan Angkatan Kerja di Jawa Tengah Agustus 2013*. Semarang: Badan Pusat Statistik Provinsi Jawa Tengah.
- BPS. 2014. *Indikator Kesejahteraan Rakyat Welfare Indicator 2014*. Semarang: Badan Pusat Statistik Provinsi Jawa Tengah.
- Han, J. and Kamber, M., 2006. *Data Mining Concepts and Techniques*. Second Edition. California: Morgan Kaufman.
- Johnson, R. A. and Wichern. D. W., 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Prentice Hall International, Inc.
- Prasetyo, E., 2012a. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Prasetyo, E., 2012b. *Fuzzy K-Nearest Neighbor In Every Class untuk Klasifikasi Data*. Seminar Nasional Teknik Informatika (Santika 2012). pp57-60.
- Prasetyo, E., 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Quinlan, J. R., 1993. *C4.5: Programs For Machine Learning*. San Mateo: Morgan Kaufmann Publisher, Inc.
- Ruggieri, S., 2002. *Efficient C4.5*. (<http://www.di.unipi.it/~ruggieri/Papers/ec45.pdf>, diakses pada tanggal 02 Maret 2015).

- Rokach, L. and Maimon, O., 2008. *Data Mining With Decision Trees: Theory and Applications*. Singapura: World Scientific Publishing Co. Pte. Ltd.
- Santosa. 2007. *Data Mining Terapan dengan MATLAB*. Yogyakarta: Graha Ilmu.
- Widodo P. P., Handayanto R. T., Herlawati. 2013. *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.
- Witten, I. H., Frank, E., Hall, M. A., 2011. *DATA MINING Practical Machine Learning Tools and Techniques*. Second Edition. California: Morgan Kaufman.